

Responses to Editor and Reviewers

I would like to express my sincerest gratitude to the editor, Olivier Talagrand, for overseeing the review of my manuscript and for providing invaluable feedback. I would also like to thank both of my reviewers, Ian Groom and anonymous, for their thorough review, commentary, feedback and kind words. This review process has helped me substantially improve the quality and clarity of the manuscript. In recognition of the editor's and reviewers' efforts, I have included them in the Acknowledgement section of my revised manuscript.

1 Responses to Reviewer 1 (Ian Groom)

1.1 General Comments

This paper presents a method for increasing the ensemble size during the assimilation step of ensemble DA. The method is interesting and the results show that it can improve performance. I have some suggestions for improving the clarity of the discussion though.

Thank you, Ian, for your kind words, thorough review and valuable feedback. Your comments, particularly regarding Section 3, have greatly enhanced the clarity of my thoughts and the discussion. I have made every effort to address your comments and hope that my efforts will bring this manuscript closer to being accepted for publication in NPG.

1.2 Major Comment

Section 3 could be improved. It attempts to show how PECE-GC could be valuable across a wide range of ensemble DA methods and gives 3 mechanisms whereby PECE-GC could improve performance. I found the whole section to be a bit too vague though. It might help to pick a few specific algorithms and show how PECE-GC could improve each one. For example, the RHF uses a piecewise-linear approximation of the likelihood function; PECE-GC improves this representation. As another example, many EnKFs assume joint Gaussianity of the state and observation vectors, and then approximate the means and cross-covariances of this distribution using an ensemble. If the prior is accurately represented then PECE-GC leads to improved estimates of these means and covariances.

Some things I found confusing about section 3: It starts with two-step ensemble DA, but not all EnKFs operate in a two-step manner. The three mechanisms seem to overlap; e.g. the likelihood function (3.1) is related to the observation operator (3.2), and to observation space ensemble statistics (3.3). Section 4 explains performance in light of these 3 mechanisms, so I also found it confusing.

Thank you for pointing out this vagueness. As you have rightly pointed out, organizing the discussion around overlapping mechanisms instead of algorithms is confusing. I have consolidated the two overlapping mechanisms (the former 3.2 and 3.3) into a single mechanism (PESE-GC influences ensemble statistics). Furthermore, I have overhauled Section 3 so that there are algorithm-by-algorithm discussions. I have also replaced the confusing two-step discussion with a discussion based on Bayes' Rule. Section 4 has also been edited to only speak of 2 mechanisms instead of 3. The overhauled Section 3 is attached in the following pages.

2.3 On the mean and variance adjustments in PESE-GC's step 3

PESE-GC requires forecast ensemble probits with zero mean and unity variance. Otherwise, the resulting virtual members will disobey the marginal distributions fitted in PESE-GC's step 1. However, because the forecast ensemble size is finite, the forecast ensemble's probits may have non-zero mean and non-unity variance. To illustrate, suppose PESE-GC is applied on 5 univariate forecast ensemble members (red crosses in Fig. 2c) and a Gaussian-tailed rank histogram distribution (Anderson, 2010) is fitted to those 5 members in PESE-GC's step 1. Applying the PPI transform (PESE-GC's step 2) results in forecast probits with mean zero and a variance of approximately 0.561. 10^7 virtual probits are then generated from these forecast probits using the CAC2020 algorithm, and the inverse PPI transform is applied to generate the virtual members (PESE-GC's step 4). The histogram-estimated PDF of the virtual members (blue curve in Fig. 2c) disagrees with the fitted (i.e., desired) Gaussian-tailed rank histogram PDF.

This problematic disagreement is resolved by adjusting the forecast probits' mean and variance to zero and unity (respectively) before generating the virtual members. Suppose the probit of the n -th forecast member for model variable i is $\phi_{i,n}$ and the pre-adjusted prior ensemble probit's sample mean and sample variance are μ and σ^2 respectively. The adjustment process is simply

$$\phi_{i,n} \leftarrow \frac{\phi_{i,n} - \mu}{\sigma} \quad \forall n = 1, 2, \dots, N_E. \quad (13)$$

The impact of this adjustment is illustrated in Fig. 2: the virtual members' histogram-estimated PDF (blue curve in Fig. 2d) now matches the Gaussian-tailed rank histogram PDF in PESE-GC's step 1.

3 Conceptual exploration of PESE-GC's EnsDA impacts

To understand the influence of PESE-GC on EnsDA, consider a joint model-observation space formulation of Bayes' rule:

$$p(\boldsymbol{\psi}|\mathbf{y}^o) \propto p(\mathbf{y}^o|\boldsymbol{\psi}) p(\boldsymbol{\psi}) \quad (14)$$

where \mathbf{y}^o is an N_y -dimensional vector of observation values,

$$\boldsymbol{\psi} \equiv \begin{bmatrix} \mathbf{x} \\ \mathbf{h}(\mathbf{x}) \end{bmatrix}, \quad (15)$$

$\mathbf{h}(\mathbf{x})$ is the observation operator, $p(\mathbf{y}^o|\boldsymbol{\psi})$ is the observation likelihood function, and $p(\boldsymbol{\psi})$ is the prior probability density function (pdf). The goal of EnsDA is to obtain a posterior ensemble that represents the posterior distribution $p(\boldsymbol{\psi}|\mathbf{y}^o)$. If PESE-GC alters $p(\mathbf{y}^o|\boldsymbol{\psi})$ and/or $p(\boldsymbol{\psi})$, then PESE-GC will influence EnsDA. As such, there are two potential mechanisms for PESE-GC to influence EnsDA: 1) alterations to the observation likelihood used by the EnsDA algorithm, and 2) alterations to the prior distribution used by the EnsDA.

This section explores the influence of PESE-GC on EnsDA through those two mechanisms. A bivariate example will be used to illustrate those mechanisms. Suppose a scalar forecast model variable x has a skewed-normal forecast distribution and

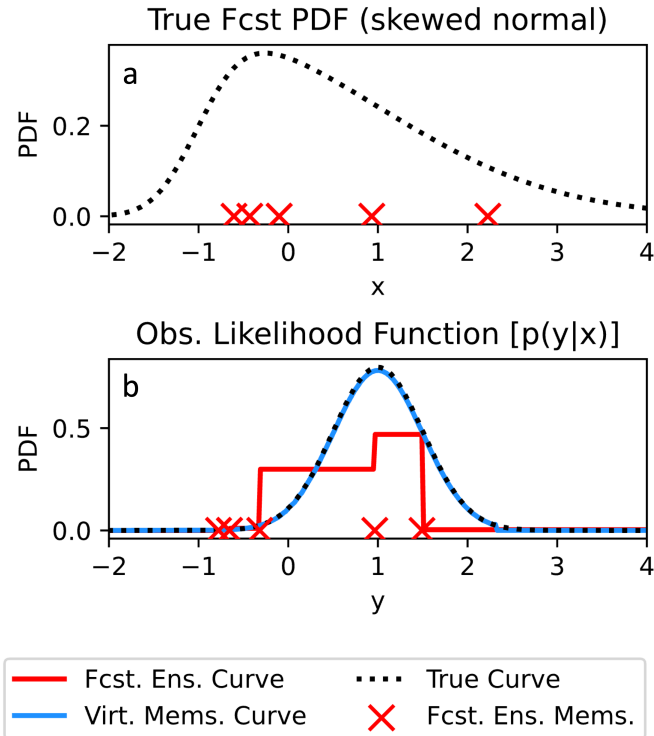


Figure 4. Plots of the true forecast PDF (a) and observation likelihood functions (b). The 5 forecast ensemble x values are indicated with red crosses in (a), and the 5 forecast ensemble y values are indicated with red crosses in (b). The red curve in (b) indicates the piecewise-approximated observation likelihood function used by a rank histogram filter that only uses the 5 forecast members. With PESE-GC, the rank histogram filter’s piecewise-approximated observation likelihood function (blue curve) is improved.

5 samples are drawn from this distribution (illustrated in Figure 4a). A signed square-root function $h(x)$

$$h(x) \equiv \text{sign}(x) \sqrt{|x|} \tag{16}$$

will be used as the observation operator, and let y denote observation values. 10,000 virtual members will be generated by PESE-GC in this bivariate example. Note that the true forecast distribution of x is the previously mentioned skewed-normal distribution (illustrated in Figure 4a), and the true forecast distribution of $h(x)$ is estimated by applying $h(x)$ on 1,000,000 samples of x drawn from the true forecast distribution of x .

3.1 Mechanism 1: PESE-GC improves ensemble-based representation of observation likelihood function

For certain EnsDA algorithms that employ ensemble-based representations of the observation likelihood function, PESE-GC can improve those representations (impact mechanism 1). Two EnsDA algorithms will be considered: 1) the Rank Histogram Filter (RHF; Anderson (2010)), and 2) the serial stochastic Ensemble Kalman Filter (serial stochastic EnKF; Burgers et al. (1998)). In the case of the RHF, an ensemble-based piecewise approximation to the observation likelihood function is used

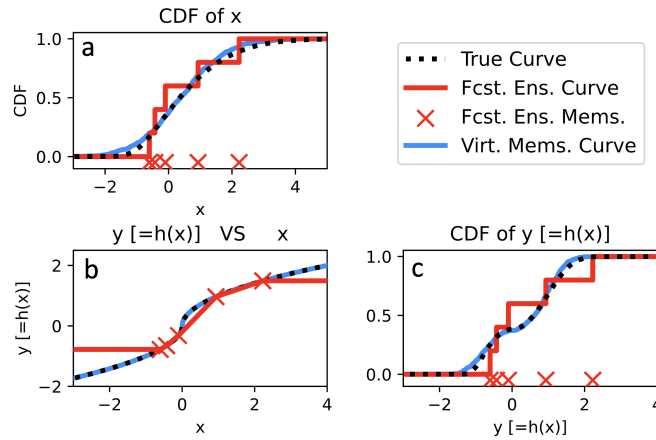


Figure 5. Bivariate example demonstrating the impacts of drawing virtual members from an informative fitted marginal (normal distribution). Panel a shows the empirical CDFs of x from the initial members and virtual members. The estimated relationships between x and y (obtained by passing the initial members and virtual members through $h(x)$) are displayed in panel b. Finally, panel c shows the empirical CDFs of the initial members and virtual members for variable y . The true CDFs and x - y relationships are plotted in panels b and c with dotted black lines.

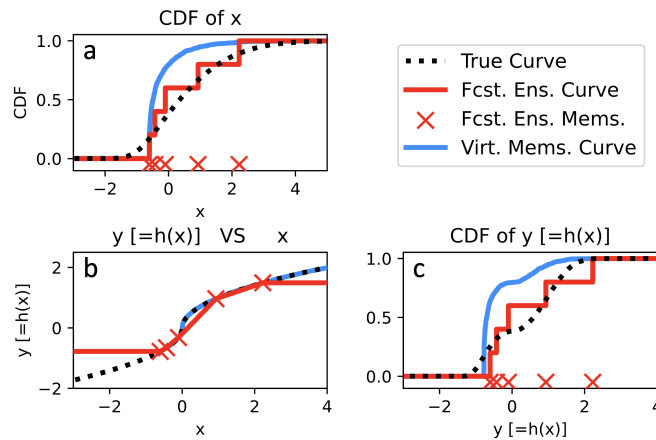


Figure 6. Bivariate example demonstrating the impacts of drawing virtual members from a misinformed fitted marginal (gamma distribution). The panels here are similar to Fig. 5.

(illustrated in Fig. 4b). The accuracy of that piecewise approximation depends on the ensemble size. When small forecast ensembles are used, that piecewise approximation is crude (red curve in Figure 4b). Since PESE-GC increases ensemble size, PESE-GC refines that piecewise representation (blue curve in Figure 4b). In the absence of other factors, this refinement will

Impact mechanism 1 also manifests for the serial stochastic EnKF. To see that, consider a situation with two observations, and recall that the serial stochastic EnKF uses random draws from a univariate Gaussian distribution to represent the likelihood function (one draw per ensemble member). For small ensembles, only a few of those random draws are made. In other words, there are sampling errors in representing the likelihood function. The ensemble statistics resulting from assimilating the first observation are thus degraded by those sampling errors. This degradation then affects the assimilation of the second observation. The assimilation of more than two observations compounds such sampling issues. Since PESE-GC increases the ensemble size, more draws from the likelihood function are made, thus suppressing sampling errors. As such, in the absence of other factors, PESE-GC will improve the performance of the serial stochastic EnKF.

Note that many EnsDA algorithms are immune to impact mechanism 1. The deterministic variants of the EnKF (Bishop et al., 2001; Whitaker and Hamill, 2002; Anderson, 2003; Tippett et al., 2003; Sakov et al., 2008) and particle filters (Gordon et al., 1993; Snyder et al., 2008; Poterjoy, 2016; van Leeuwen et al., 2019) are immune to impact mechanism 1 because no ensemble-based representation of the likelihood function is used. Also, the stochastic EnKF that assimilates all observations simultaneously (all-at-once stochastic EnKF) is immune to this effect – the chain of events described in the previous paragraph will not occur for the all-at-once stochastic EnKF. Finally, EnsDA systems that run ensembles of variational data assimilation methods (e.g., the ECMWF) are also immune to impact mechanism 1. This immunity is due to the same reason as the all-at-once stochastic EnKF.

3.2 Mechanism 2: PESE-GC influences ensemble statistics

If the user specifies appropriate marginals in PESE-GC’s stage 1, then PESE-GC will improve the ensemble statistics used by EnsDA algorithms (mechanism 2). This will be illustrated using the bivariate 5-member example discussed near the start of Section 3. Suppose the user knows that the prior marginal distribution of x is close to Gaussian. Thus, a Gaussian distribution is fitted to the forecast x values in PESE-GC’s stage 1 (Section 2.2; Fig. 5a). Applying PESE-GC generates 10,000 virtual members. Figure 5 indicates that the virtual members have better ensemble statistics than the forecast ensemble. This improvement in ensemble statistics is a combination of 1) the user selecting an appropriate marginal for x , and 2) the nonlinear $h(x)$ is evaluated more often (i.e., better sampled) with a large ensemble. As such, if the user specifies appropriate marginals in PESE-GC’s stage 1, PESE-GC will likely improve the performance of EnsDA algorithms.

An important caveat is that if the user selects misinformed marginal distributions, then PESE-GC may degrade the ensemble statistics used by EnsDA algorithms. To illustrate, suppose the user fits a shifted gamma distribution (Cheng and Amin, 1983) to the 5 forecast x values. This distribution has three parameters: shape, scale and location. Since only 5 forecast values are used to fit three parameters, the fitted parameters’ sampling errors are severe. Applying PESE-GC with this badly estimated shifted gamma distribution results in virtual ensemble statistics that are worse than those of 5 forecast x values (Fig. 6). In such situations, the performance of EnsDA will likely be degraded by PESE-GC. As such, the selection of marginals to use with PESE-GC must be done with care.

In the absence of knowledge about the model variables’ prior marginal distribution, users can use non-parametric marginal distributions with PESE-GC. Such distributions include the Gaussian-tailed rank histogram distribution (Anderson, 2010) and

230 kernel distributions (Anderson and Anderson, 1999). Choosing such distributions may not improve the model-space ensemble statistics. However, because PESE-GC increases the number of observation operator evaluations, the observation-space ensemble statistics and the ensemble covariance/copula between observation and model quantities can be improved.

Note that when linear observations are assimilated, PESE-GC with Gaussian marginals will not change the performance of deterministic variants of the EnKF (Bishop et al., 2001; Whitaker and Hamill, 2002; Anderson, 2003; Tippett et al., 2003; 235 Sakov et al., 2008). This is because the expanded ensemble will have the same joint-space mean and covariance as the forecast ensemble.

4 Tests with Lorenz 1996 model

4.1 Setup of experiments

This section explores the impacts of PESE-GC on the performance of EnsDA using perfect model Observing System Simulation 240 tion Experiments (OSSEs) with the Lorenz 1996 model (L96 model; Lorenz (2006)). The Data Assimilation Research Testbed (DART; <https://github.com/NCAR/DART>; Anderson et al. (2009)) is used in this exploration and PESE-GC has been implemented into DART.

The L96 model uses 40 variables (i.e., 40 grid points in a ring), a forcing parameter value of 8 (i.e., $F = 8$), and a time-step of 0.05 L96 time units. The L96 time unit is henceforth referred to as τ . Forward time integration of the model is done via the 245 Runge-Kutta fourth-order integration scheme. Every OSSE experiment runs for 5,500 cycles. Initial nature run states and the initial ensemble members are drawn from the L96's climatology.

Note that the 0.05τ time step has been arbitrarily taken to be either 1 hour (Anderson, 2019, 2023) or 6 hours (Lorenz, 2006; Kurosawa and Poterjoy, 2021). To avoid potential confusion, this paper will avoid converting L96 time units to hours. In other words, all results will be displayed and discussed in terms of L96 time units.

250 In all experiments, there are 40 observations. Their observation locations are fixed throughout this study. Supposing that the model grid points have locations $0.025m$, $m = 1, 2, \dots, 40$, each site location is a random draw from a uniform distribution between 0 and 1.

PESE-GC's impacts are examined using EnsDA experiments that are conducted with four N_{ES} , five cycling intervals, three post-PESE-GC ensemble sizes, three observation types, four EnsDA algorithms, and with and without PESE-GC, a total of 255 $4 \times 5 \times 3 \times 3 \times 4 \times 2 = 1440$ configurations. The four N_{ES} are 10, 20, 40, and 80, and the five cycling intervals are 0.05τ , 0.10τ , 0.15τ , 0.30τ , and 0.60τ .

The PESE-GC-expanded ensemble sizes are specified in terms of factors: 5, 10 and 20 times the forecast ensemble size. For example, if $N_E = 10$, a PESE-GC factor of 10 means that the expanded ensemble has 100 members (i.e., $N_V = 90$ virtual members are created).

1.3 Minor Comments

1.3.1 'Gaussian copula' is in the name of the method, but the concept of a Gaussian copula is not explained until section 5.

Thanks for catching that. I have added the following paragraph in Section 2 (right after describing the 4-step PESE-GC procedure).

"Note that this four-stage assumes that the multivariate forecast distribution is Gaussian in probit space. This assumption arises from the use of a Gaussian resampling algorithm (the CAC2020 algorithm) to generate virtual probits. This assumption is equivalent to assuming that the multivariate forecast distribution has a Gaussian copula. As such, this four-stage procedure is called "Probit-space Ensemble Size Expansion for Gaussian Copulas"."

1.3.2 Line 15/16: I struggle to see how 'small forecast ensembles result in limited representation of observation likelihood functions' is a distinct concept from sampling errors. PECE-GC seems to me to be a way to mitigate certain kinds of sampling errors.

Thanks for catching this lack of clarity in the first lines of the introduction. Your interpretation is correct – PESE-GC mitigates certain kinds of sampling errors. I have updated the discussion in Section 3 to reflect that thinking (see attached pages). The updated introduction now reads:

"Geophysical forecast models are often computationally expensive to run. As a result, geophysical ensemble data assimilation (EnsDA) typically uses <100 forecast ensemble members. Such small forecast ensemble sizes result in sampling errors that degrade the performance of EnsDA. As such, low-cost methods that introduce additional ensemble members (henceforth, "virtual members") to the original forecast members (henceforth, "forecast members") have the potential to improve EnsDA."

1.3.3 Lines 33/34: It might be more accurate to say that ensemble modulation preserves the first two moments of the ensemble, not that it assumes Gaussian statistics.

You are correct. I have adjusted the paragraph in question. That paragraph now reads:

"Ensemble modulation (Bishop and Hodyss, 2009, 2011; Bishop et al., 2017; Kotsuki and Bishop, 2022; Wang et al., 2021) is the fourth type of ensemble expansion method. Such methods expand ensembles by combining a localization matrix with the original ensembles' perturbations (see Bishop and Hodyss (2009) for details). Though the expanded ensemble possesses the same mean and variance as the original ensemble, the expanded ensemble's kurtosis can be much larger than the original ensemble's (see the Supplement). In other words, the expanded ensemble's kurtosis is likely biased. If nonlinear observation operators are applied on the expanded ensemble, this kurtosis bias will result in biased expanded ensemble observation statistics (personal communication with Craig Bishop and Lili Lei)."

1.3.4 Line 218: I think 0.05 model time units is usually interpreted as 6 hours, not 1 hour.

Thanks for pointing this out. I noticed some inconsistencies in the literature regarding the interpretation of those time units. To prevent potential reader confusion, I have removed all conversions from L96 time units to hours. The relevant lines now read:

"The L96 model uses 40 variables (i.e., 40 grid points in a ring), a forcing parameter value of 8 (i.e., $F = 8$), and a time-step of 0.05 L96 time units. The L96 time unit is henceforth referred to as τ . Forward time integration of the model is done via the Runge-Kutta fourth-order integration scheme. Every OSSE experiment runs for 5,500 cycles. Initial nature run states and the initial ensemble members are drawn from the L96's climatology.

Note that the 0.05τ time step has been arbitrarily taken to be either 1 hour (Anderson, 2019, 2023) or 6 hours (Lorenz, 2006; Kurosawa and Poterjoy, 2021). To avoid potential confusion, this paper will avoid converting L96

time units to hours. In other words, all results will be displayed and discussed in terms of L96 time units."

Figures 7, 8 and 9 have also been modified to use τ instead of hours. I have also replaced all uses of hours with τ .

1.3.5 Line 259: The fifth-order GC function is rational, not polynomial.

Thank you for catching this. The relevant line now reads:

*"The Gaspari-Cohn fifth-order **rational** function..."*

1.3.6 Line 272: Missing citation for RTPS.

Thank you for catching that. The citation is now added to the relevant line.

*"...or a relaxation-to-posterior-spread (RTPS; **Whitaker and Hamill (2012)**)..."*

1.3.7 I do not understand why PECE-GC has any impact on the performance of the stochastic EnKF with linear obs. With linear obs the stochastic EnKF update is entirely controlled by the ensemble covariance B, and PECE-GC does not change B (except in the PR configuration).

Thank you for asking this question. This question helped me to understand why PESE-GC influenced the stochastic EnKF. PESE-GC specifically influences the performance of the serial stochastic EnKF. I have added the following paragraph to Section 3.1 to explain why.

*"**Impact mechanism 1 also manifests for the serial stochastic EnKF. To see that, consider a situation with two observations, and recall that the serial stochastic EnKF uses random draws from a univariate Gaussian distribution to represent the likelihood function (one draw per ensemble member). For***

small ensembles, only a few of those random draws are made. In other words, there are sampling errors in representing the likelihood function. The ensemble statistics resulting from assimilating the first observation are thus degraded by those sampling errors. This degradation then affects the assimilation of the second observation. The assimilation of more than two observations compounds such sampling issues. Since PESE-GC increases the ensemble size, more draws from the likelihood function are made, thus suppressing sampling errors. As such, in the absence of other factors, PESE-GC will improve the performance of the serial stochastic EnKF."

In contrast, I do not expect PESE-GC to have any impact on the stochastic EnKF that simultaneously assimilates all observations in one shot. I have added the following lines to Section 3.1 (around line 185) to explain why.

"Also, the stochastic EnKF that assimilates all observations simultaneously (all-at-once stochastic EnKF) is immune to this effect – the chain of events described in the previous paragraph will not occur for the all-at-once stochastic EnKF."

1.3.8 Line 402: 1,000,00 should probably be 1,000,000?

Thank you for catching that typographical error. It has been fixed.

1.3.9 The GC part of PECE-GC seems to only have been used in the PR configuration, is that right? The EAKF, EnKF, and RHF all use PECE without the GC part?

You are correct. To be clear, the EAKF, EnKF and RHF experiments effectively used the CAC2020 algorithm – this is a result of using PESE-GC with Gaussian marginals. It is only in the PR experiments that I am actually employing non-Gaussian marginals. To clarify matters, I have added the following lines to the paragraph starting in Line 256:

"For each EnsDA algorithm, only 1 set of marginals are used with PESE-GC. When PESE-GC is used with the EAKF, EnKF or RHF algorithm, Gaussian

marginals are selected for all 40 model variables. PESE-GC with Gaussian marginals is identical to the CAC2020 algorithm. In other words, for the EAKF, EnKF and RHF experiments, the virtual ensemble members follow multivariate Gaussian distributions. For the PR algorithm, the Gaussian-tailed rank histogram is selected as the marginal for every one of the 40 model variables. This means the PR experiments' virtual ensemble members follow multivariate non-Gaussian distributions. Future work can investigate the impacts of using PESE-GC with Gaussian-tailed rank histograms (or kernel density estimates) with the EAKF, EnKF and RHF."

2 Responses to Reviewer 2

2.1 General Comments

This manuscript nicely proposed a method to expand the ensemble size in probit space. Unlike commonly used methods for ensemble size expansion, this PSES-GC ensemble size expansion method can generate virtual members from any multivariate forecast distribution with a Gaussian copula, by using the users' knowledge of the marginal distributions of forecast model variables. This PSES-GC method has been tested with tremendous scenarios, using the L96 model. Results show that PSES-GC can improve the performances of ensemble-based data assimilation methods when the actual ensemble size is small, or the marginal distributions improves the forecast model variable statistics, or the rank histogram filter is used with non-parametric priors. Expanding the ensemble size has advantages to ensemble-based DA in many aspects, while this PSES-GC is an interesting one. The manuscript is well written and pleasant to read. I have several specific comments as below.

Thank you for your kind words and thorough review. Your comments have significantly improved this manuscript. I have made every effort to address your comments and hope that my efforts will bring this manuscript closer to being accepted for publication in NPG.

2.2 Minor Comments

2.2.1 L28, I'd like to bring up two references here. Sun et al. (2022, 2024) use "analog" ensemble for paleoclimate data assimilation. The paleoclimate data assimilation could be a potential application for the ensemble size expansion, since very limited ensemble members are available.

Sun, H., L. Lei, Z. Liu, L. Ning, and Z.-M. Tan, 2024: A hybrid gain analog offline EnKF for paleoclimate data assimilation. J. Adv. Model. Earth Syst., 16, e2022MS003414, doi: <https://doi.org/10.1029/2022MS003414>

Sun, H., L. Lei, Z. Liu, L. Ning, and Z.-M. Tan, 2022: An analog offline EnKF for paleoclimate data assimilation. J. Adv. Model. Earth Syst., 14, e2021MS002674, doi: <https://doi.org/10.1029/2021MS002674>

Thank you for your suggestions. I have added those references to the relevant paragraph in my introduction. That paragraph now reads:

"A third type of ensemble expansion method is to search a historical catalog for forecast states similar to the current forecast or observations (Van den Dool, 1994; Tippett and Delsole, 2013; Monache et al., 2013; Wan and Van Der Merwe, 2000; Grooms, 2021; Sun et al., 2022). The virtual members resulting from this search have flow-dependent statistics. Though such methods are historically expensive to employ, ongoing research may render them affordable in the near future (e.g., Sun et al. (2024))."

2.2.2 L33-34, "However, ensemble modulation assumes that forecast uncertainties have Gaussian statistics." This statement is not right clear to me. Could you explain a little bit why the modulation assumes Gaussian statistics? The modulation just applies localization, not really an ensemble size expansion.

Thank you for pointing out this issue. Reviewer 1 (Ian Groom) also pointed out this issue. I agree with both of you that ensemble modulation does not assume Gaussian statistics.

I agree with you that modulation does apply localization. However, the resulting modulated ensemble is much larger than the original unlocalized ensemble. As such, I consider ensemble modulation as an ensemble expansion technique.

The relevant paragraph in the introduction has been edited to remove the mistaken assumption about Gaussian statistics (see below).

"Ensemble modulation (Bishop and Hodyss, 2009, 2011; Bishop et al., 2017; Kotsuki and Bishop, 2022; Wang et al., 2021) is the fourth type of ensemble expansion method. Such methods expand ensembles by combining a localization matrix with the original ensembles' perturbations (see Bishop and Hodyss (2009) for details). Though the expanded ensemble possesses the same mean and variance as the original ensemble, the expanded ensemble's kurtosis can be much larger than the original ensemble's (see the Supplement). In other words, the expanded ensemble's kurtosis is likely biased. If nonlinear observation operators are applied on the expanded ensemble, this kurtosis bias will result in biased expanded ensemble observation statistics (personal communication with Craig Bishop and Lili Lei)."

2.2.3 L89, it would be nice to have Eq. (6) of Chan et al. (2023) here. Then the reader can immediately see the difference between Eq. (6) of Chan et al. (2023) and the \mathbf{W} in this manuscript.

Thank you for your suggestion. I have added a simplified form of that equation. The relevant paragraph now reads:

"Note that this study's (and the CAC2020's) \mathbf{W} differs from that of Chan et al. (2023) (henceforth, "CAC2023"). This is because the CAC2023's \mathbf{W} [defined in Eq. (6) of CAC2023] generates virtual members with undesirable non-Gaussian properties. The CAC2023's \mathbf{W} can be written as

$$\text{CAC2023's } \mathbf{W}: \quad W_{ij} \equiv \delta_{i,j} - \frac{1}{N_V} \quad (1)$$

where $\delta_{i,j}$ is the Kronecker-delta.

*To illustrate the issue with the CAC2023's **W**, suppose 5 forecast members are drawn from a standard normal distribution..."*

2.2.4 Section 2.2 and Figure 3, it is not straightforward to see how the cross-variable relationships are handled? Are the covariances of model state variables naturally conserved by performing PPI for each kind of state variable?

Again at I132, does the adjustment conserve the covariance in model space?

Great questions! I am handling these questions together because they are related. If Gaussian marginals are chosen for PESE-GC, the PPIs and the adjustments to probit-space mean and covariance will not change the covariance in model space (i.e., the covariance is conserved). However, if non-Gaussian marginals are chosen for PESE-GC, the covariances in model space will likely be changed. The way the covariances change likely depends on what marginals are chosen and the distribution fitting procedure (e.g., moment-matching versus maximum likelihood). This subtlety may have interesting effects (e.g., may increase covariance matrix rank) and deserves future investigation.

I have modified the fifth paragraph of my concluding section to reflect the open question that you have posed (see below). Thank you for this comment!

*"Another general area for future work is to develop the PESE-GC algorithm further. First, given the importance of localization in practical EnsDA, future work can and should explore inserting localization into PESE-GC. Second, the validity of PESE-GC's Gaussian copula assumption can be assessed in the context of geophysical modelling and forecasting. If the Gaussian copula assumption is inappropriate, then non-parametric methods to generate virtual probits can be explored. **Third**, methods to detect the usage of misinformed parametric marginal distribution families deserve exploration. One possible detection method is to employ hypothesis testing on the marginal distributions. For example, if Gaussian distributions are selected for PESE-GC, then the Shapiro-Wilk test can be applied on the forecast ensemble to determine if the selection*

is misinformed (e.g., Kurosawa and Poterjoy (2023)). Finally, the use of non-Gaussian marginals with PESE-GC may alter ensemble covariances between model variables. This possibility deserves future investigation."

2.2.5 L160-165, it is not quite clear the difference between the components A and C. Are they both about $p(y|x)$?

Thank you for your question. Component A is about $p(y|x)$, and component C is about the prior statistics of the simulated observations (i.e., related to $p(x)$).

Note that have I rewritten Section 3 in response to the concerns of Reviewer 1 (Ian Groom). Instead of three components, there are now two components – one for the observation likelihood, the other for the prior distribution. I have attached the re-written Section 3 in the following pages.

2.3 On the mean and variance adjustments in PESE-GC's step 3

PESE-GC requires forecast ensemble probits with zero mean and unity variance. Otherwise, the resulting virtual members will disobey the marginal distributions fitted in PESE-GC's step 1. However, because the forecast ensemble size is finite, the forecast ensemble's probits may have non-zero mean and non-unity variance. To illustrate, suppose PESE-GC is applied on 5 univariate forecast ensemble members (red crosses in Fig. 2c) and a Gaussian-tailed rank histogram distribution (Anderson, 2010) is fitted to those 5 members in PESE-GC's step 1. Applying the PPI transform (PESE-GC's step 2) results in forecast probits with mean zero and a variance of approximately 0.561. 10^7 virtual probits are then generated from these forecast probits using the CAC2020 algorithm, and the inverse PPI transform is applied to generate the virtual members (PESE-GC's step 4). The histogram-estimated PDF of the virtual members (blue curve in Fig. 2c) disagrees with the fitted (i.e., desired) Gaussian-tailed rank histogram PDF.

This problematic disagreement is resolved by adjusting the forecast probits' mean and variance to zero and unity (respectively) before generating the virtual members. Suppose the probit of the n -th forecast member for model variable i is $\phi_{i,n}$ and the pre-adjusted prior ensemble probit's sample mean and sample variance are μ and σ^2 respectively. The adjustment process is simply

$$\phi_{i,n} \leftarrow \frac{\phi_{i,n} - \mu}{\sigma} \quad \forall n = 1, 2, \dots, N_E. \quad (13)$$

The impact of this adjustment is illustrated in Fig. 2: the virtual members' histogram-estimated PDF (blue curve in Fig. 2d) now matches the Gaussian-tailed rank histogram PDF in PESE-GC's step 1.

3 Conceptual exploration of PESE-GC's EnsDA impacts

To understand the influence of PESE-GC on EnsDA, consider a joint model-observation space formulation of Bayes' rule:

$$p(\boldsymbol{\psi}|\mathbf{y}^o) \propto p(\mathbf{y}^o|\boldsymbol{\psi}) p(\boldsymbol{\psi}) \quad (14)$$

where \mathbf{y}^o is an N_y -dimensional vector of observation values,

$$\boldsymbol{\psi} \equiv \begin{bmatrix} \mathbf{x} \\ \mathbf{h}(\mathbf{x}) \end{bmatrix}, \quad (15)$$

$\mathbf{h}(\mathbf{x})$ is the observation operator, $p(\mathbf{y}^o|\boldsymbol{\psi})$ is the observation likelihood function, and $p(\boldsymbol{\psi})$ is the prior probability density function (pdf). The goal of EnsDA is to obtain a posterior ensemble that represents the posterior distribution $p(\boldsymbol{\psi}|\mathbf{y}^o)$. If PESE-GC alters $p(\mathbf{y}^o|\boldsymbol{\psi})$ and/or $p(\boldsymbol{\psi})$, then PESE-GC will influence EnsDA. As such, there are two potential mechanisms for PESE-GC to influence EnsDA: 1) alterations to the observation likelihood used by the EnsDA algorithm, and 2) alterations to the prior distribution used by the EnsDA.

This section explores the influence of PESE-GC on EnsDA through those two mechanisms. A bivariate example will be used to illustrate those mechanisms. Suppose a scalar forecast model variable x has a skewed-normal forecast distribution and

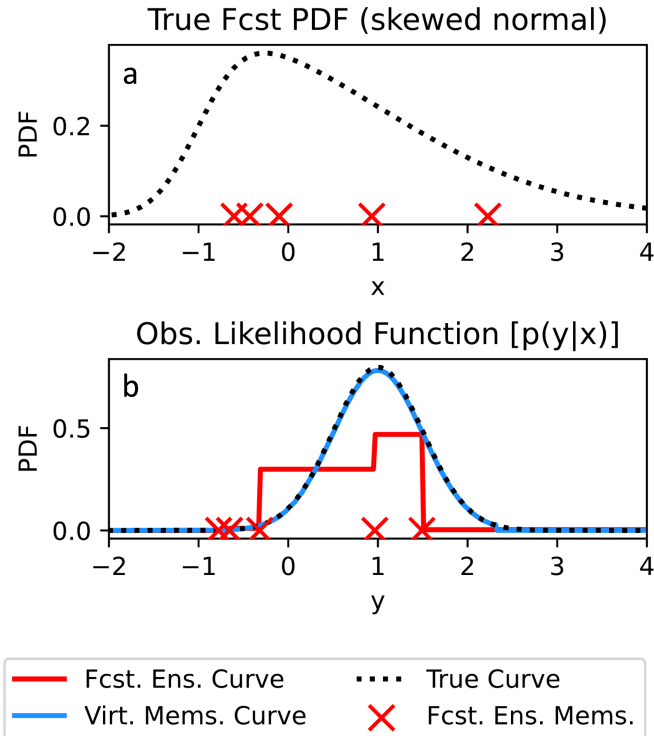


Figure 4. Plots of the true forecast PDF (a) and observation likelihood functions (b). The 5 forecast ensemble x values are indicated with red crosses in (a), and the 5 forecast ensemble y values are indicated with red crosses in (b). The red curve in (b) indicates the piecewise-approximated observation likelihood function used by a rank histogram filter that only uses the 5 forecast members. With PESE-GC, the rank histogram filter’s piecewise-approximated observation likelihood function (blue curve) is improved.

5 samples are drawn from this distribution (illustrated in Figure 4a). A signed square-root function $h(x)$

$$h(x) \equiv \text{sign}(x) \sqrt{|x|} \tag{16}$$

will be used as the observation operator, and let y denote observation values. 10,000 virtual members will be generated by PESE-GC in this bivariate example. Note that the true forecast distribution of x is the previously mentioned skewed-normal distribution (illustrated in Figure 4a), and the true forecast distribution of $h(x)$ is estimated by applying $h(x)$ on 1,000,000 samples of x drawn from the true forecast distribution of x .

3.1 Mechanism 1: PESE-GC improves ensemble-based representation of observation likelihood function

For certain EnsDA algorithms that employ ensemble-based representations of the observation likelihood function, PESE-GC can improve those representations (impact mechanism 1). Two EnsDA algorithms will be considered: 1) the Rank Histogram Filter (RHF; Anderson (2010)), and 2) the serial stochastic Ensemble Kalman Filter (serial stochastic EnKF; Burgers et al. (1998)). In the case of the RHF, an ensemble-based piecewise approximation to the observation likelihood function is used

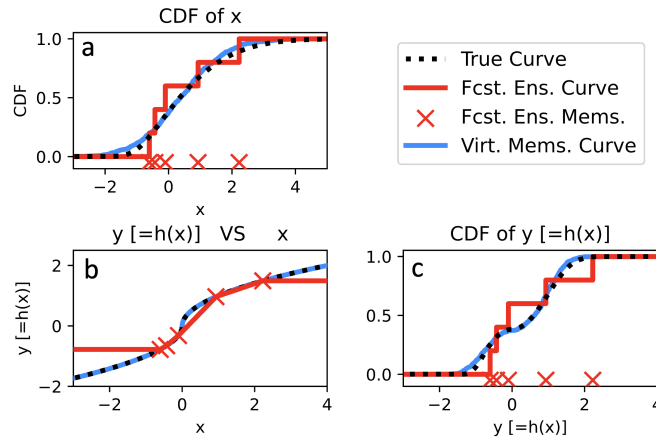


Figure 5. Bivariate example demonstrating the impacts of drawing virtual members from an informative fitted marginal (normal distribution). Panel a shows the empirical CDFs of x from the initial members and virtual members. The estimated relationships between x and y (obtained by passing the initial members and virtual members through $h(x)$) are displayed in panel b. Finally, panel c shows the empirical CDFs of the initial members and virtual members for variable y . The true CDFs and x - y relationships are plotted in panels b and c with dotted black lines.

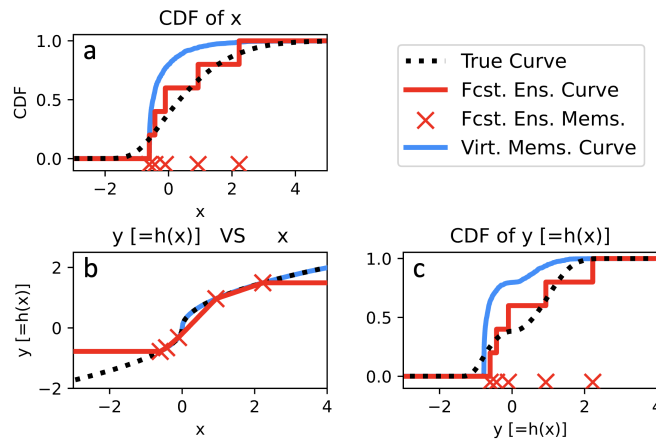


Figure 6. Bivariate example demonstrating the impacts of drawing virtual members from a misinformed fitted marginal (gamma distribution). The panels here are similar to Fig. 5.

(illustrated in Fig. 4b). The accuracy of that piecewise approximation depends on the ensemble size. When small forecast ensembles are used, that piecewise approximation is crude (red curve in Figure 4b). Since PESE-GC increases ensemble size, PESE-GC refines that piecewise representation (blue curve in Figure 4b). In the absence of other factors, this refinement will

Impact mechanism 1 also manifests for the serial stochastic EnKF. To see that, consider a situation with two observations, and recall that the serial stochastic EnKF uses random draws from a univariate Gaussian distribution to represent the likelihood function (one draw per ensemble member). For small ensembles, only a few of those random draws are made. In other words, there are sampling errors in representing the likelihood function. The ensemble statistics resulting from assimilating the first observation are thus degraded by those sampling errors. This degradation then affects the assimilation of the second observation. The assimilation of more than two observations compounds such sampling issues. Since PESE-GC increases the ensemble size, more draws from the likelihood function are made, thus suppressing sampling errors. As such, in the absence of other factors, PESE-GC will improve the performance of the serial stochastic EnKF.

Note that many EnsDA algorithms are immune to impact mechanism 1. The deterministic variants of the EnKF (Bishop et al., 2001; Whitaker and Hamill, 2002; Anderson, 2003; Tippett et al., 2003; Sakov et al., 2008) and particle filters (Gordon et al., 1993; Snyder et al., 2008; Poterjoy, 2016; van Leeuwen et al., 2019) are immune to impact mechanism 1 because no ensemble-based representation of the likelihood function is used. Also, the stochastic EnKF that assimilates all observations simultaneously (all-at-once stochastic EnKF) is immune to this effect – the chain of events described in the previous paragraph will not occur for the all-at-once stochastic EnKF. Finally, EnsDA systems that run ensembles of variational data assimilation methods (e.g., the ECMWF) are also immune to impact mechanism 1. This immunity is due to the same reason as the all-at-once stochastic EnKF.

3.2 Mechanism 2: PESE-GC influences ensemble statistics

If the user specifies appropriate marginals in PESE-GC’s stage 1, then PESE-GC will improve the ensemble statistics used by EnsDA algorithms (mechanism 2). This will be illustrated using the bivariate 5-member example discussed near the start of Section 3. Suppose the user knows that the prior marginal distribution of x is close to Gaussian. Thus, a Gaussian distribution is fitted to the forecast x values in PESE-GC’s stage 1 (Section 2.2; Fig. 5a). Applying PESE-GC generates 10,000 virtual members. Figure 5 indicates that the virtual members have better ensemble statistics than the forecast ensemble. This improvement in ensemble statistics is a combination of 1) the user selecting an appropriate marginal for x , and 2) the nonlinear $h(x)$ is evaluated more often (i.e., better sampled) with a large ensemble. As such, if the user specifies appropriate marginals in PESE-GC’s stage 1, PESE-GC will likely improve the performance of EnsDA algorithms.

An important caveat is that if the user selects misinformed marginal distributions, then PESE-GC may degrade the ensemble statistics used by EnsDA algorithms. To illustrate, suppose the user fits a shifted gamma distribution (Cheng and Amin, 1983) to the 5 forecast x values. This distribution has three parameters: shape, scale and location. Since only 5 forecast values are used to fit three parameters, the fitted parameters’ sampling errors are severe. Applying PESE-GC with this badly estimated shifted gamma distribution results in virtual ensemble statistics that are worse than those of 5 forecast x values (Fig. 6). In such situations, the performance of EnsDA will likely be degraded by PESE-GC. As such, the selection of marginals to use with PESE-GC must be done with care.

In the absence of knowledge about the model variables’ prior marginal distribution, users can use non-parametric marginal distributions with PESE-GC. Such distributions include the Gaussian-tailed rank histogram distribution (Anderson, 2010) and

230 kernel distributions (Anderson and Anderson, 1999). Choosing such distributions may not improve the model-space ensemble statistics. However, because PESE-GC increases the number of observation operator evaluations, the observation-space ensemble statistics and the ensemble covariance/copula between observation and model quantities can be improved.

Note that when linear observations are assimilated, PESE-GC with Gaussian marginals will not change the performance of deterministic variants of the EnKF (Bishop et al., 2001; Whitaker and Hamill, 2002; Anderson, 2003; Tippett et al., 2003; 235 Sakov et al., 2008). This is because the expanded ensemble will have the same joint-space mean and covariance as the forecast ensemble.

4 Tests with Lorenz 1996 model

4.1 Setup of experiments

This section explores the impacts of PESE-GC on the performance of EnsDA using perfect model Observing System Simulation 240 tion Experiments (OSSEs) with the Lorenz 1996 model (L96 model; Lorenz (2006)). The Data Assimilation Research Testbed (DART; <https://github.com/NCAR/DART>; Anderson et al. (2009)) is used in this exploration and PESE-GC has been implemented into DART.

The L96 model uses 40 variables (i.e., 40 grid points in a ring), a forcing parameter value of 8 (i.e., $F = 8$), and a time-step of 0.05 L96 time units. The L96 time unit is henceforth referred to as τ . Forward time integration of the model is done via the 245 Runge-Kutta fourth-order integration scheme. Every OSSE experiment runs for 5,500 cycles. Initial nature run states and the initial ensemble members are drawn from the L96's climatology.

Note that the 0.05τ time step has been arbitrarily taken to be either 1 hour (Anderson, 2019, 2023) or 6 hours (Lorenz, 2006; Kurosawa and Poterjoy, 2021). To avoid potential confusion, this paper will avoid converting L96 time units to hours. In other words, all results will be displayed and discussed in terms of L96 time units.

250 In all experiments, there are 40 observations. Their observation locations are fixed throughout this study. Supposing that the model grid points have locations $0.025m$, $m = 1, 2, \dots, 40$, each site location is a random draw from a uniform distribution between 0 and 1.

PESE-GC's impacts are examined using EnsDA experiments that are conducted with four N_{ES} , five cycling intervals, three post-PESE-GC ensemble sizes, three observation types, four EnsDA algorithms, and with and without PESE-GC, a total of 255 $4 \times 5 \times 3 \times 3 \times 4 \times 2 = 1440$ configurations. The four N_{ES} are 10, 20, 40, and 80, and the five cycling intervals are 0.05τ , 0.10τ , 0.15τ , 0.30τ , and 0.60τ .

The PESE-GC-expanded ensemble sizes are specified in terms of factors: 5, 10 and 20 times the forecast ensemble size. For example, if $N_E = 10$, a PESE-GC factor of 10 means that the expanded ensemble has 100 members (i.e., $N_V = 90$ virtual members are created).

2.2.6 L239, any explanation for the choice of the error variance of 0.25 and 16 for the SQRT and SQUARE observations, given 1.0 for IDEN?

Thank you for spotting this omission. Those values are taken from Anderson (2020) and I have updated the relevant sentence to treat that omission. The relevant sentence now reads:

Following Anderson (2020), the σ^2 values for the IDEN, SQRT and SQUARE observations are set to 1.0, 0.25, and 16, respectively.

2.2.7 L272, the reference for RTPS is missing.

Thank you for catching that error. I have furnished the reference for RTPS (Whitaker and Hamill, 2012) in the manuscript.

2.2.8 Section 4.4, here mechanisms 1-3 are used. It would be better to consistently use components A, B and C as in previous discussions.

Thank you for your suggestion. I have replaced the three components with two mechanisms in response to Reviewer 1's comments. The discussion in Section 4 now uses the same terminology as in Section 3.

2.2.9 L323-327, I don't quite understand here. I thought it should be more Gaussian if in probit space.

Good question. The issue here is that Gaussian marginals are chosen for the PPIs used for PESE-GC in the EAKF, EnKF and RHF experiments. The Gaussian PPI transform can be shown to be a rescaling and recentering procedure. In other words, if the ensemble distribution in native model space is non-Gaussian, then the resulting distribution in probit space will also be non-Gaussian. In other words, the bad performance here is simply because inappropriate marginals are chosen for the PPI (the ensemble is probably following

a non-Gaussian distribution).

2.2.10 L340-342, I am not convinced here. The RHF can be seen as a non-Gaussian filter.

Thank you for your concern! I agree with you that the RHF is a non-Gaussian filter. However, a major difference between the RHF experiments and the PR experiments is in the the marginals chosen for PESE-GC – Gaussian for RHF and Gaussian-tailed rank histogram for PR. The issue here is PESE-GC with Gaussian marginals are used even though the true model-space forecast distribution is non-Gaussian. In other words, the expanded ensemble’s statistics are worse than the forecast ensemble’s statistics. PESE-GC with non-Gaussian marginals (e.g., Gaussian-tailed rank histogram marginals) is likely more appropriate in this situation. Even though the RHF is a non-Gaussian filter, if the ensemble statistics themselves are degraded, then the RHF’s performance will suffer.

2.2.11 It is nice to virtually increase the ensemble size. But it would be nicer to discuss the computational cost for the DA process, with increased ensemble sizes.

Thank you for your suggestion. I have added the following paragraphs to Section 5.

"It is also important to discuss the impacts of PESE-GC on the EnsDA process (i.e., the forecast step and analysis step). Since the virtual members are deleted before running forecast models (Figure 1), PESE-GC does not change the forecast step’s computational cost. However, PESE-GC will increase the computational cost of the analysis step because 1) the number of observation operator calls scales linearly with the ensemble size, and 2) the EnsDA filter algorithm’s (e.g., the EAKF algorithm) computational complexity scales with the ensemble size. Supposing the computational cost of the EnsDA filter algorithm scales linearly with ensemble size, then the computational cost of the EnsDA analysis step scales linearly with the number of ensemble members created by PESE-GC.

However, the increase in the computational cost associated with PESE-GC is likely far more affordable than running a larger forecast ensemble. This is because the computational cost of the forecast step often accounts for ~ 80% of the overall computational cost of the EnsDA process and the analysis step accounts for the remaining ~ 20%. Consider the situation where PESE-GC quintuples the ensemble size. The overall computational cost of the EnsDA process will only increase by ~ 80%. In contrast, if the forecast ensemble size is quintupled, then the overall computational cost of the EnsDA process increases by ~ 400%. As such, PESE-GC is an alluring alternative to increasing the forecast ensemble size."

3 Responses to the Editor

The two referees have evaluated positively the paper, and recommend acceptance, subject to minor revisions. Referee 1 (who has let his name known, and is Ian Grooms) has a major comment about Section 3 of the paper, which he finds a ‘bit too vague’, and for which he would like clearer explanations. He has in addition a number of minor, mostly editorial, comments. Referee 2 has also a number of minor comments, which have to do with both scientific and editorial aspects. I as Editor have also a number of comments (I mention that there seems to be a shift of a few units between the line numbers below and the ones mentioned by the referees).

Thank you for summarizing the two reviewer’s contributions and for your thorough review. I greatly appreciate the effort that you and the reviewers have spent on this manuscript. Your comments have significantly improved this manuscript and I have made every effort to address them. I hope that my efforts will bring this manuscript closer to being accepted for publication in NPG.

3.1 Conceptual comments

3.1.1 I wonder in particular what the virtues of the extended PESE-GC ensemble are. I am sure these virtues are described in the previous publications of the author (in particular Chan et al., 2020, and Chan, 2022), but it could be useful to say a little more in the present paper for readers who are new to that method. ... In particular does the CAC2020 algorithm (subsection 2.1) introduce additional properties of interest ? Additional explanations on those points, even succinct, could be useful (this aspect may have to do with Referee 1’s major comment).

Thank you for pointing out this deficiency. My previous publications do not discuss PESE-GC at all. Instead, my previous publications discuss an efficient and embarrassingly parallel Gaussian resampling scheme. The Gaussian nature of

that scheme is the scheme's primary weakness. This manuscript extends that Gaussian resampling scheme to handle non-Gaussian distributions. I have repurposed Section 2.1.5 to describe the properties of that Gaussian resampling scheme (the CAC2020 algorithm). The following paragraphs have been added to Section 2.1.5 as part of the repurposing.

"The CAC2020 algorithm also produces expanded ensembles with same ensemble means and ensemble covariances as the forecast ensembles. In other words, the rank of the expanded ensemble's covariance matrix is the same as that of the forecast ensemble. Future work can explore ways to incorporate localization into the expanded ensemble's covariance matrix.

Furthermore, the CAC2020 algorithm always generates Gaussian-distributed virtual members: even if the actual forecast distribution is highly non-Gaussian, the virtual members' distribution will still be Gaussian. The CAC2020 algorithm thus degrades the ensemble statistics in situations where the forecast distribution is non-Gaussian. This degradation limits the usefulness of the CAC2020 algorithm for situations with non-Gaussian forecast distributions.

Note that, except for the mean and covariance, the expanded ensemble's central moments (i.e., higher-order moments; e.g., skewness) likely differ from the forecast ensemble's. More specifically, the expanded ensemble's central moments will be closer to those of Gaussian distributions (e.g., zero skewness) than the forecast ensemble's central moments. This is because the virtual members are effectively drawn from a Gaussian distribution. If the forecast distribution is indeed a Gaussian distribution, then the expanded ensemble likely has better moments than the forecast ensemble."

3.1.2 From what I understand, specific properties of the extended PESE-GC ensemble are that it has the same mean and covariance matrix as the forecast ensemble, that it preserves through PPI the marginal distributions of the forecast model variables, and that it is numerically very economical.

Thank you for your feedback. I would like to highlight that PESE-GC specifically conserves the mean and covariance in probit space (i.e., the transformed space). If the marginals chosen are Gaussian, then PESE-GC does conserve

the mean and covariance in the original space (i.e., untransformed). However, if non-Gaussian marginals are chosen, then it is not immediately clear what PESE-GC does to the mean and covariance in the original space. I have added a sentence to the fifth paragraph of my concluding section to point out this area of future work (see below).

"Another general area for future work is to develop the PESE-GC algorithm... Finally, the use of non-Gaussian marginals with PESE-GC may alter ensemble covariances between model variables. This possibility deserves future investigation."

3.1.3 Would any other ensemble expansion method that had the same properties (maybe there is not any) be as useful?

I am not aware of any other ensemble expansion methods with the numerical economy and generality of the PESE-GC algorithm. To highlight this, I have added a sentence to the end of the sixth paragraph in Section 1 (see below).

"The shortcomings of existing ensemble expansion methods motivate... To the author's knowledge, no other ensemble expansion method simultaneously possesses the same efficiency, scalability, generality and flow-dependency as PESE-GC."

3.1.4 My understanding is that the CAC2020 algorithm is implemented between steps 3 and 4 of the PESE-GC procedure (II. 122-126). Say it there. And, if I am mistaken, additional appropriate information will be useful.

Thank you for pointing out my lack of clarity. I have added two sentences stating the usage of the CAC2020 algorithm in stage 3 of the PESE-GC procedure. The first sentence occurs right after the in-text description of the 4 stages of PESE-GC:

"To be clear, the CAC2020 algorithm is implemented and used in stage 3."

The second sentence occurs inside the caption of Figure 3:

"...The details of these stages are described in Section 2.2. **Note that the CAC2020 algorithm is applied in stage 3 (i.e., between panels c and d).**"

3.1.5 Step 3 of PESE-GC and subsection 2.3. Is the need to adjust the mean and variance of each variable to 0 and 1 due only to the finiteness of the forecast ensemble ? If yes, say it explicitly. If no, explain more clearly.

Thank you for your suggestion. Your understanding is correct. I have added a sentence saying that the finiteness of the forecast ensemble is the source of the problem. That sentence is in the first paragraph of Section 2.3.

"PESE-GC requires forecast ensemble probits with zero mean and unity variance. Otherwise, the resulting virtual members will disobey the marginal distributions fitted in PESE-GC's step 1. **However, because the forecast ensemble size is finite, the forecast ensemble's probits may have non-zero mean and non-unity variance.** To illustrate..."

3.2 Editorial comments

3.2.1 Eq. (2), denominator on the rhs $N_E \rightarrow N_V$

Thank you for catching my typographical error. I have fixed that error.

3.2.2 Eq. (4). What does Chol(.) exactly mean here ? The Cholesky decomposition of a symmetric matrix C is defined by $C = U U^T$, where U is a triangular matrix. What is Chol(C) in Eq. (4) ? Either one of those two triangular matrices, or what ?

Good question. I am specifically using the lower-triangular matrix is produced by the Cholesky decomposition (i.e., U). I have corrected my description of the Chol(*) operator:

"... $\text{Chol}(\cdot)$ represents Cholesky decomposition (which produces a lower-triangular matrix), ..."

3.2.3 LI. 81-82. I presume the $\omega_{i,j}$ are mutually independent ? Say it explicitly.

You are correct. I have added a statement of that mutual independence after Eq. (7):

"where all $\omega_{i,j}$ s are identically and independently distributed (i.i.d.) samples drawn from the standard normal distribution. In other words, all $\omega_{i,j}$ s are mutually independent."

3.2.4 L. 109. Since no particular meaning would apparently be given to $F_i(x_j)$ for $i \neq j$, one index i is sufficient (the double index may actually be confusing). I suggest to write $F_i(z)$, where z is a dummy real argument.

Thank you for your suggestion. However, there is no index j in the equation you've referenced – there are two occurrences of the same index i . I deliberately put two i indices on $F_i(x_i)$ to highlight that the CDF varies with the model variable in question (i is the model variable index). I have added the following sentence to the paragraph before Eq. (10):

"...and ϕ_i represents the i -th model probit. The double appearance of index i in $F_i(x_i)$ is deliberate – the CDF varies with the chosen model variable. Note that..."

3.2.5 End of caption of Fig.5. dashed ... lines → dotted ...

The correction has been made. The last sentence in Fig. 5's caption now reads

"The true CDFs and x - y relationships are plotted in panels b and c with dotted black lines."

3.2.6 L. 241, values of variance of the observation error σ^2 do not make much sense without some appropriate scale of reference (for instance, the climatological variance of the solutions of the L96 model).

Thank you for pointing this out. Reviewer 2 also made a similar comment. Those σ^2 values are lifted from Anderson (2020). The relevant sentence now reads:

Following Anderson (2020), the σ^2 values for the IDEN, SQRT and SQUARE observations are set to 1.0, 0.25, and 16, respectively.

3.2.7 L. 386 and further below. Gaussian variable $G(x_1; -1, 2)$. The argument x_1 is here useless. And why not use the established notation $N(-1, 4)$ for gaussian variables ($N(\text{expectation, variance})$) ?

Thank you for pointing this out. However, my use of $G(x_1; -1, 2)$ instead of $N(-1, 2)$ is deliberate: $N(-1, 2)$ represents the Gaussian distribution, but $G(x_1; -1, 2)$ represents the **PDF** of the Gaussian distribution. To illustrate, consider the equation

$$p(x) = N(-1, 2).$$

When (inexperienced) readers look at that equation, they may get confused and ask "where did the x go in the rhs?".

If I am saying " x is drawn from $N(-1, 2)$ ", then the notation is (to my understanding) $x \sim N(-1, 2)$. However, because the equation is describing the PDF of x_1 , not sampling x_1 , then $G(x_1; -1, 2)$ is more appropriate. As such, I would like to keep my notation as is.

3.2.8 L. 79, I_N is the ... identity matrix

I have made the correction (an→the). The sentence now reads:

"... \mathbf{I}_{N_E} is the $N_E \times N_E$ identity matrix"

**3.2.9 L. 237, ... the model variable interpolated to location ...
Which kind of interpolation ?**

Thank you for catching that. Linear interpolation is used. The sentence now reads:

"... x_{l_k} is the model variable linearly interpolated to location l_k ..."

References

- Anderson, J. L.: A nonlinear rank regression method for ensemble Kalman filter data assimilation, <https://doi.org/10.1175/MWR-D-18-0448.1>, 2019.
- Anderson, J. L.: A marginal adjustment rank histogram filter for non-Gaussian ensemble data assimilation, *Monthly Weather Review*, 148, <https://doi.org/10.1175/MWR-D-19-0307.1>, 2020.
- Anderson, J. L.: A Quantile-Conserving Ensemble Filter Framework. Part II: Regression of Observation Increments in a Probit and Probability Integral Transformed Space, *Monthly Weather Review*, 2023.
- Bishop, C. H. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere, *Tellus, Series A: Dynamic Meteorology and Oceanography*, 61 A, <https://doi.org/10.1111/j.1600-0870.2008.00372.x>, 2009.
- Bishop, C. H. and Hodyss, D.: Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation, *Monthly Weather Review*, 139, <https://doi.org/10.1175/2010MWR3403.1>, 2011.
- Bishop, C. H., Whitaker, J. S., and Lei, L.: Gain form of the ensemble transform Kalman Filter and its relevance to satellite data assimilation with model space ensemble covariance localization, *Monthly Weather Review*, 145, <https://doi.org/10.1175/MWR-D-17-0102.1>, 2017.
- Chan, M.-Y., Chen, X., and Anderson, J. L.: The potential benefits of handling mixture statistics via a bi-Gaussian EnKF: tests with all-sky satellite infrared radiances, *Journal of Advances in Modeling Earth Systems*, <https://doi.org/10.1029/2022MS003357>, 2023.
- Grooms, I.: Analog ensemble data assimilation and a method for constructing analogs with variational autoencoders, *Quarterly Journal of the Royal Meteorological Society*, 147, <https://doi.org/10.1002/qj.3910>, 2021.
- Kotsuki, S. and Bishop, C. H.: Implementing Hybrid Background Error Covariance into the LETKF with Attenuation-Based Localization: Experiments with a Simplified AGCM, *Monthly Weather Review*, 150, <https://doi.org/10.1175/MWR-D-21-0174.1>, 2022.

- Kurosawa, K. and Poterjoy, J.: Data Assimilation Challenges Posed by Non-linear Operators: A Comparative Study of Ensemble and Variational Filters and Smoothers, *Monthly Weather Review*, 149, 2369–2389, <https://doi.org/10.1175/MWR-D-20-0368.1>, 2021.
- Kurosawa, K. and Poterjoy, J.: A Statistical Hypothesis Testing Strategy for Adaptively Blending Particle Filters and Ensemble Kalman Filters for Data Assimilation, *Monthly Weather Review*, 151, 105–125, <https://doi.org/10.1175/MWR-D-22-0108.1>, 2023.
- Lorenz, E. N.: Predictability—a problem partly solved, in: *Predictability of Weather and Climate*, <https://doi.org/10.1017/CBO9780511617652.004>, 2006.
- Monache, L. D., Anthony Eckel, F., Rife, D. L., Nagarajan, B., and Searight, K.: Probabilistic weather prediction with an analog ensemble, *Monthly Weather Review*, 141, <https://doi.org/10.1175/MWR-D-12-00281.1>, 2013.
- Sun, H., Lei, L., Liu, Z., Ning, L., and Tan, Z. M.: An Analog Offline EnKF for Paleoclimate Data Assimilation, *Journal of Advances in Modeling Earth Systems*, 14, <https://doi.org/10.1029/2021MS002674>, 2022.
- Sun, H., Lei, L., Liu, Z., Ning, L., and Tan, Z. M.: A Hybrid Gain Analog Offline EnKF for Paleoclimate Data Assimilation, *Journal of Advances in Modeling Earth Systems*, 16, <https://doi.org/10.1029/2022MS003414>, 2024.
- Tippett, M. K. and Delsole, T.: Constructed analogs and linear regression, *Monthly Weather Review*, 141, <https://doi.org/10.1175/MWR-D-12-00223.1>, 2013.
- Van den Dool, H. M.: Searching for analogues, how long must we wait?, *Tellus A*, 46, <https://doi.org/10.1034/j.1600-0870.1994.t01-2-00006.x>, 1994.
- Wan, E. and Van Der Merwe, R.: The unscented Kalman filter for nonlinear estimation, in: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pp. 153–158, IEEE, <https://doi.org/10.1109/ASSPCC.2000.882463>, 2000.
- Wang, X., Chipilski, H. G., Bishop, C. H., Satterfield, E., Baker, N., and Whitaker, J. S.: A multiscale local gain form ensemble transform Kalman

filter (MLGETKF), Monthly Weather Review, 149, <https://doi.org/10.1175/MWR-D-20-0290.1>, 2021.

Whitaker, J. S. and Hamill, T. M.: Evaluating methods to account for system errors in ensemble data assimilation, Monthly Weather Review, 140, <https://doi.org/10.1175/MWR-D-11-00276.1>, 2012.